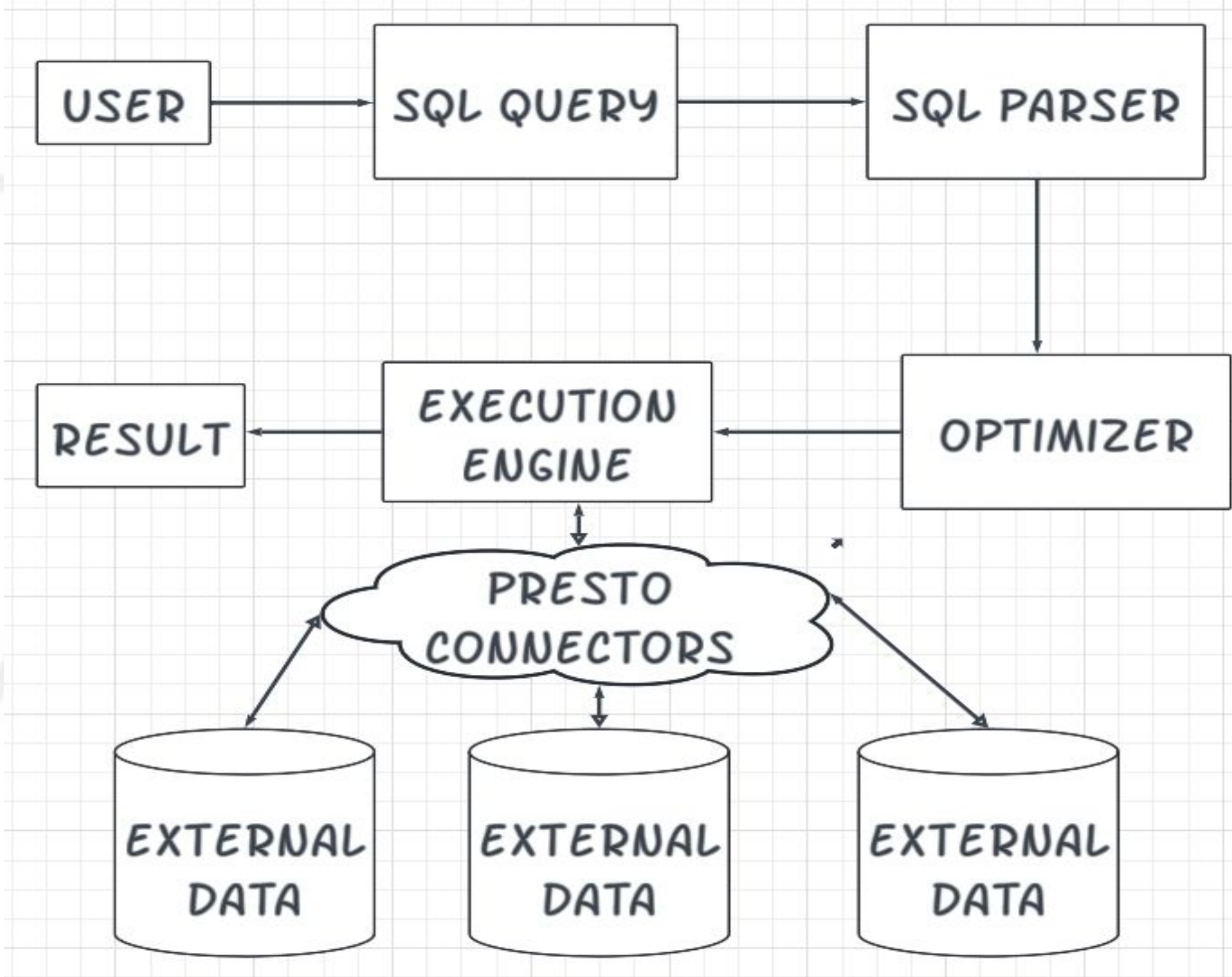


Project Overview

Invented by Meta, continue to be developed by Meta, Bytedance, Uber, IBM and more.

Our tasks are to improve the optimizer with new rules like null rows removal in aggregation and constant value optimization. Also, we want to document Presto's HTTP endpoints using the OpenAPI standard, enabling automatic generation of documentation and in the future, Presto clients.



Goals and Milestones 1

- Submit the first PR for documentation;
- Improve the optimizer to remove null rows before aggregation query;
- Submit the first PR "AddTest to TestPage"
- Complete the Presto Guides and identify the areas to contribute the internship

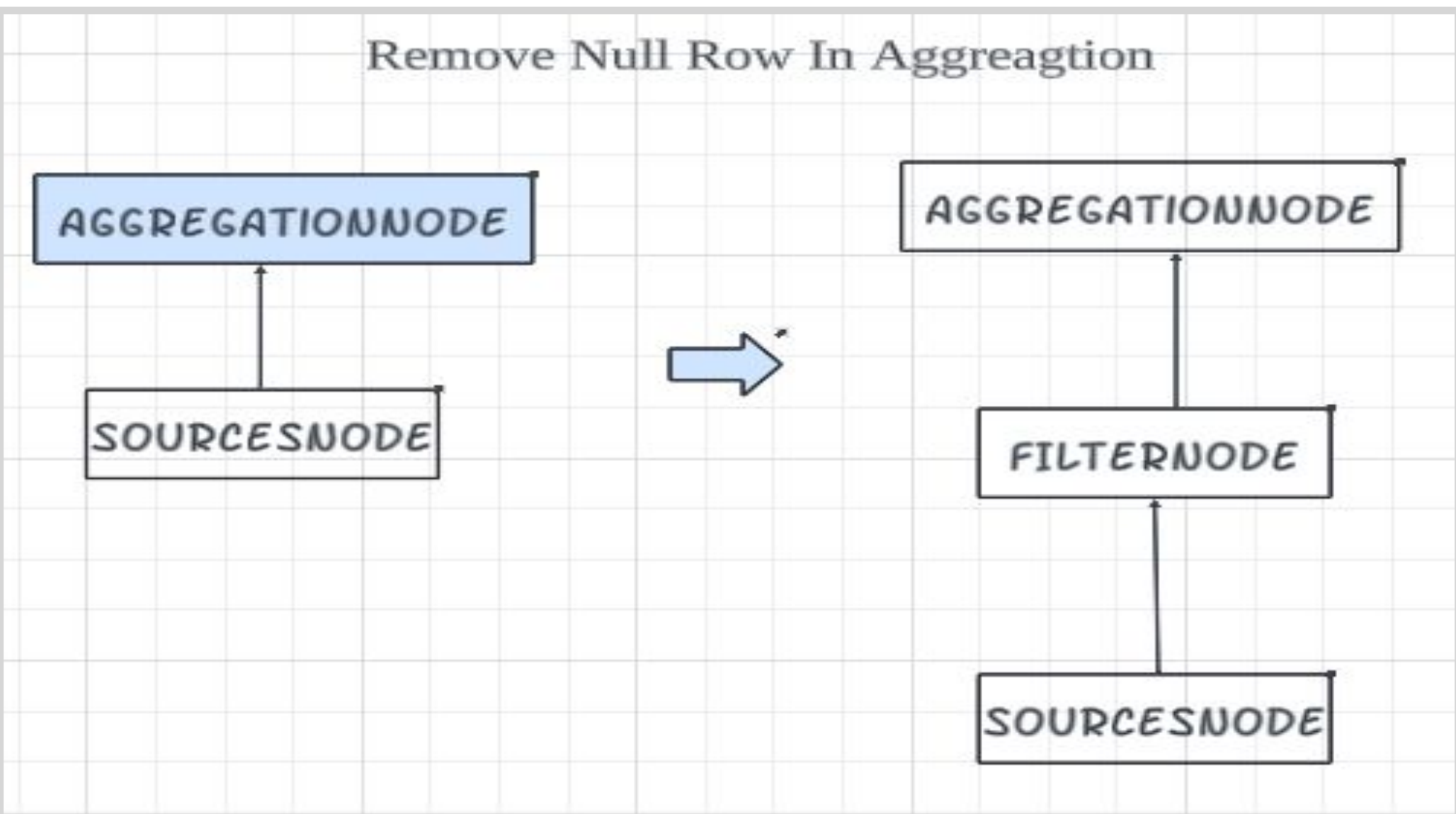
Goals and Milestones 2

- Improve the optimizer to evaluate Project in simple query
- Improve the optimizer to replace semi join operation with filter in simple query
- Understand and document Presto's endpoints using OpenAPI
- Use OpenAPI to document Presto's HTTP endpoints

Highlights and Accomplishments SubProject 1

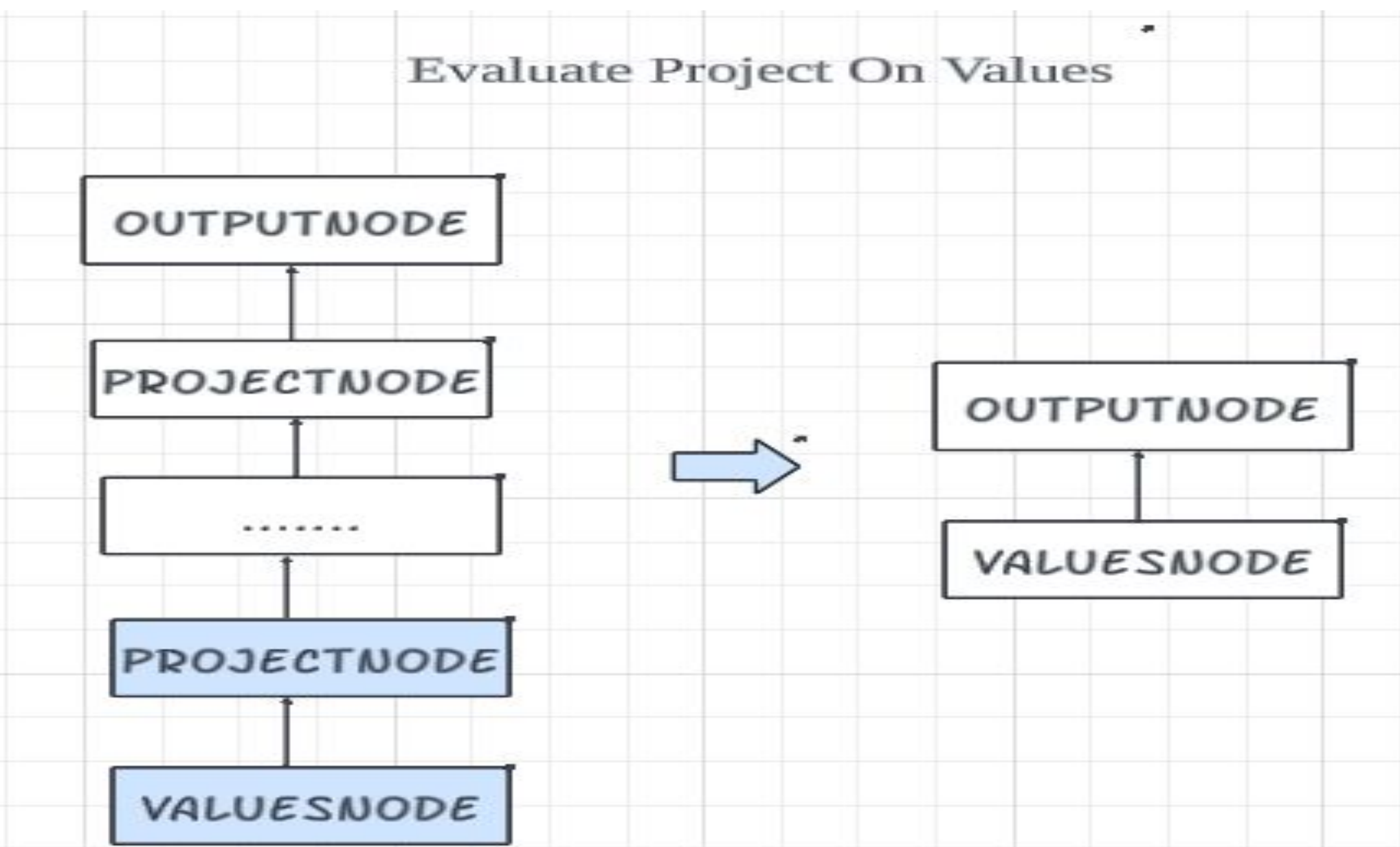
```

SELECT avg(a) FROM (VALUES 1, NULL, 2, NULL) t(a)
SELECT avg(a) FROM (VALUES 1, NULL, 2, NULL) WHERE a is not NULL
  
```



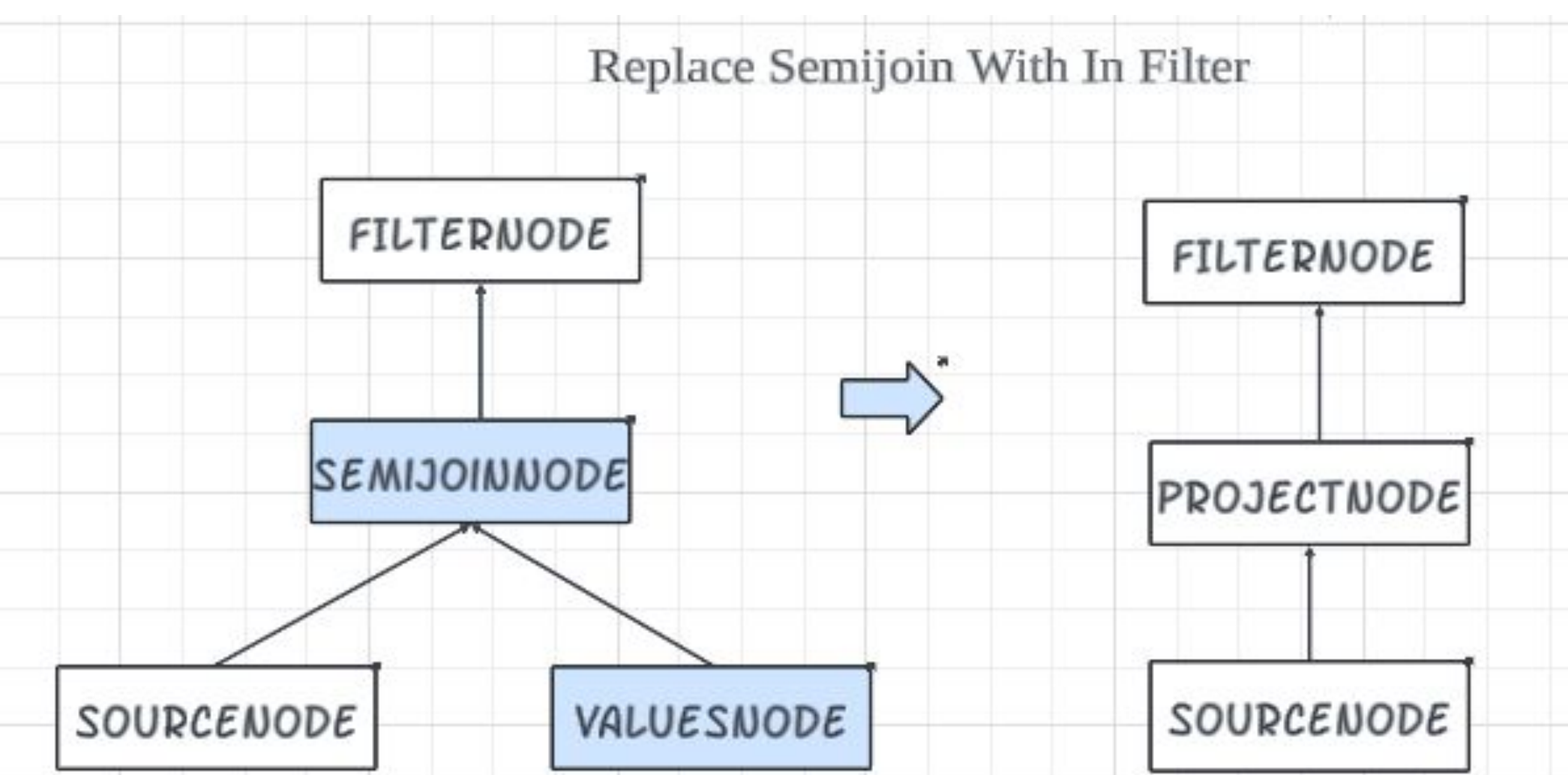
```

SELECT COUNT(*) FROM (VALUES 1,2,3,4,5)
SELECT 5
  
```



```

SELECT * FROM (VALUES (1,5),(2,6),(3,7))
t(k,v) WHERE k in (VALUES 1,2);
  
```



Highlights and Accomplishments SubProject 2

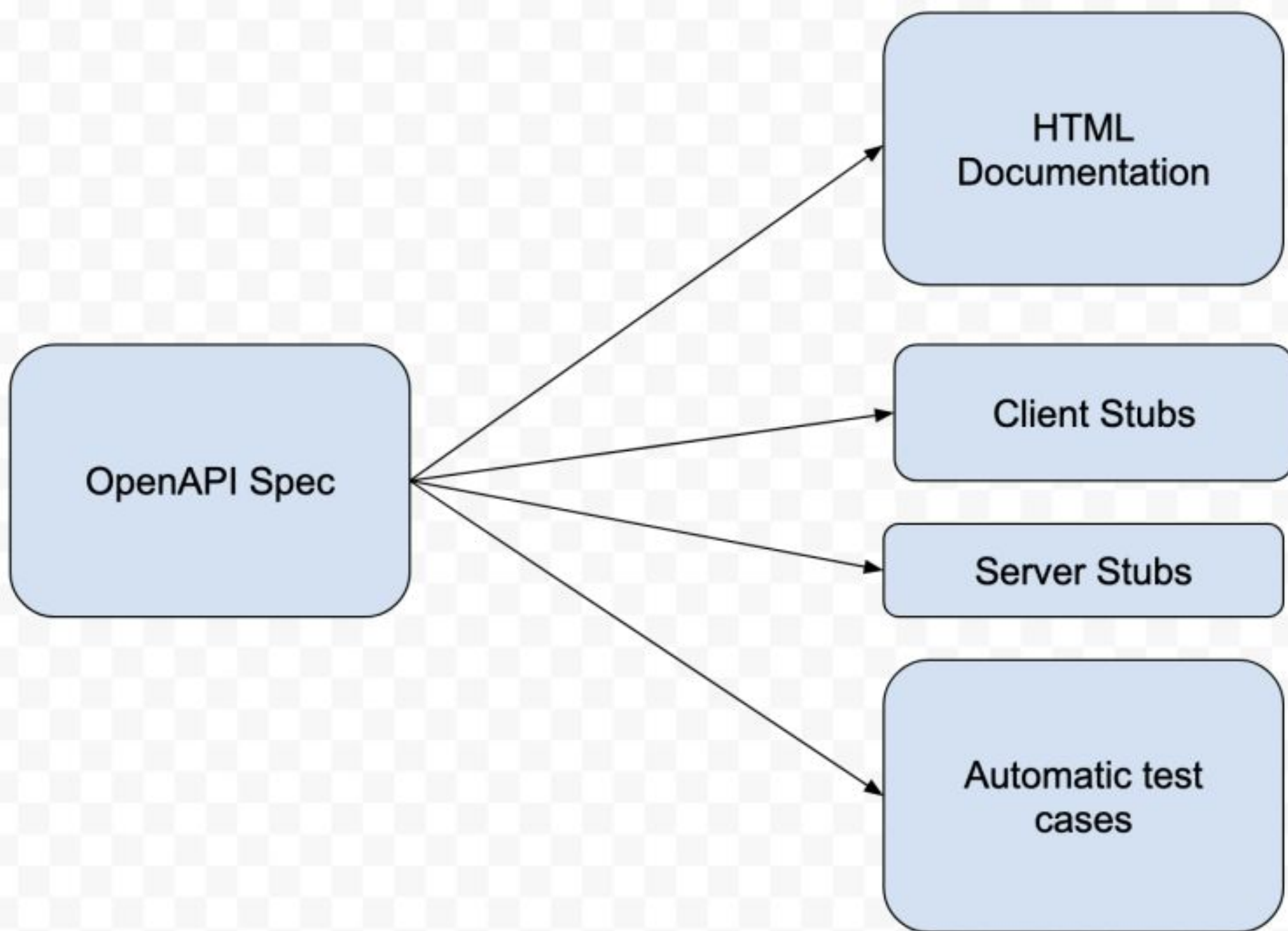
Mastering OpenAPI:

- Gained a comprehensive understanding of OpenAPI and familiarized with Swagger Specification and grammar

What is OpenAPI?

The OpenAPI Specification defines a standard, language-agnostic interface to RESTful APIs which allows both humans and computers to discover and understand the capabilities of a service without access to source code, documentation, or through network traffic inspection.

The various outputs that can be generated from an OpenAPI specification:



API Documentation and Contributions:

- Created a new module, "presto-openapi," to house the OpenAPI specification..
- Documented Presto's query-related endpoints in various files such as query.yaml, cluster.yaml, and queued.yaml. For example, cluster.yaml refers to Cluster status information which includes cluster metadata and getting information about running worker and memory allocation; query resource are for submitting queries to a cluster to execute.
- Generated HTML documentation from OpenAPI specifications for easy reference.

Open Source Outcomes

- Removing null row before aggregation;
- Evaluating project on values in simple query saves query time in query plan execution
- Rewriting a semijoin can have considerable performance improvement.
- Uniform doc across all endpoints ensures a consistent experience for developers working with the API.
- Automatically generated client stubs in various programming languages which help to simplify the the process of integrating Presto into other systems.

Future Work

- Improve constant folding using same logic as execution engine.
- Further study on DB query optimization theory and practice. Hope to contribute in the future.
- Ensure that all existing endpoints are fully documented, and look for others not been covered.
- Use the OpenAPI specifications to generate tests that ensure the API behaves as documented.